图书情报工作
第 65 卷 第 13 期 2021 年 7 月

强弱连接对学科引证知识扩散动态链路预测的影响研究*

■ 岳增慧[1] 许海云[2] 赵敏[3]

[1] 济宁医学院医学信息工程学院 日照 276826 [2] 山东理工大学管理学院 淄博 255000

[3] 济宁医学院 日照 276826

摘 要：[目的/意义]强弱连接是影响学科引证知识扩散动态链路预测的重要因素之一。学科知识扩散强弱引证连接相互协同、相互影响，共同促进了学科间的知识交流、融合与创新。学科引证知识扩散动态链路预测中强弱连接效应的探索，可为强弱连接理论应用场景的拓展，学科引证知识扩散行为微观演化规律的揭示以及动态链路预测算法指标的评价、设计与优化提供理论与实践参考。[方法/过程]依托内外协同的思路理念，构建一种外部网络结构调控与内部微观演化机理剖析相结合的动态链路预测强弱连接效应探测方法，分别从学科引证知识关联权重调节、连边失效触发以及强弱连接模体分析三个维度，对基于共同邻居相似性的学科引证知识扩散动态链路预测中的强弱连接效应问题进行探讨。[结果/结论]强连接在学科引证知识扩散网络演化及动态链路预测过程中扮演着更加重要的角色；链路预测中的强弱连接现象不仅与学科引证关联权重有关，还会受到共同邻居数目以及网络微观模体结构的影响；知识宿学科的吸纳融合能力相对于知识源学科的溢出辐射能力来说，在新连边衍生过程中的主导地位更加突出。

关键词：学科引证知识扩散 强弱连接 权重调节 连边失效 三元组模体 动态链路预测

分类号：G250

DOI：10.13266/j.issn.0252-3116.2021.13.007

1974 年，美国社会学家 M. S. Granovetter 提出了弱连接理论[1]。在传统社会，与最亲密的人接触是一种稳定但传播范围有限的社会认知，称为“强连接”现象，强连接关系代表着行动者之间具有高度的互动。与此同时，存在一种更为广泛但相对浅显的社会认知，即“弱连接”现象，弱连接虽然不如强连接那样坚固，却具有低成本和高效能的传播特点[1-2]。强弱连接在网络衍生、信息传播扩散等方面扮演着重要的角色。强连接维系了网络社团内部的高凝聚性互动关联，双方行动者可以从强连接中优先、轻易获得更强有力的支撑，比弱连接更能有效地传递复杂信息和隐性知识[3]，然而却可能带来信息冗余、闭塞等问题。相对于强连接，弱连接虽不稳定却连接范围广泛，行动者资源异质化特征明显，能够在不同社团间传递非重复的新知识和新信息，弱连接在知识迁移、组织创新、信息流动等方面都发挥了重要的作用[4-7]。学科引证知识扩散网络中节点之间的相互作用强度和类型亦具有一定差异，即存在学科引证强连接和弱连接。学科间的引证频次（即关联权重）较大的连边属于强连接，较小的连边属于弱连接。强弱引证连接相互协同、相互影响，共同促进了学科间的知识交流、融合与创新。

学科引证知识关联随着时间不断演化，如何基于现时网络信息动态准确预测学科之间未来的引证链路，即学科引证知识扩散动态链路预测问题，是明晰学科知识流变趋势、辅助学科知识管理决策的有力支撑。对未来可能产生的连边进行预测的核心是对网络演化规律的把握，网络微观结构（尤其是强弱连接模式）的变化对网络演化起到了至关重要的作用。模体（mo-

* 本文系国家自然科学基金青年项目“学科知识扩散规律及动力学机制研究”（项目编号：71704063）和国家自然科学基金青年项目“基于科学-技术主题关联分析的创新演化路径识别方法研究”（项目编号：71704170）研究成果之一。

作者简介：岳增慧（ORCID：0000-0002-9192-9071），副教授，博士，E-mail：yzh66123@126.com；许海云（ORCID：0000-0002-7453-3331），教授，博士；赵敏（ORCID：0000-0002-6190-1481），教授，博士。

收稿日期：2020-10-10 修回日期：2021-02-20 本文起止页码：66-76 本文责任编辑：杜杏叶

岳增慧，许海云，赵敏. 强弱连接对学科引证知识扩散动态链路预测的影响研究[J]. 图书情报工作,2021,65(13):66 - 76.

tif),即网络中频繁出现的局部连接模式[8],作为一种重要的网络微观结构,其演化特性可以客观地揭示网络结构特征的变化[9]。如果能够准确掌握强弱连接微观结构的演化规律,分析微观模体的变化趋势,便能够更加有效地预测网络的整体演化方向,进而促进链路预测算法的设计与改进。有研究发现权重较大的强连接在链路预测中起到了较大的作用,即链路预测中存在强连接效应[10];而有的研究则发现权重较小的弱连接扮演着更为重要的角色,即链路预测中存在弱连接效应[11]。根据前期研究发现,在对学科引证知识扩散演化网络进行预测的过程中,有些算法指标考虑权重后预测效果得到提升,有些则不然,即存在一定程度的强弱连接效应[12],说明强弱连接是影响学科引证知识扩散动态链路预测的重要因素之一。学科引证知识扩散动态链路预测中强弱连接效应的探索,可为强弱连接理论应用场景的拓展,学科引证知识扩散行为微观演化规律的揭示以及动态链路预测算法指标的评价、设计与优化提供理论与实践参考。

1 相关研究回顾

目前,国内外学者针对链路预测中的强弱连接问题进行了许多探索。T. Murata 等基于可通过网络拓扑结构和连接权重更好地估计节点之间邻近度的假设,介绍了一种利用网络加权邻近度来进行链路预测的改进方法,并在密集社交网络上证实了该方法的有效性[10]。L. Y. Lü 等在加权网络链路预测指标的基础上,引入调节参数,探讨了网络中强连接和弱连接对于链路预测的作用,并利用模体分析方法给出了半定量解释[11]。H. Liu 等提出了一种基于共同邻居节点中心度和弱连接的链路预测模型,实证研究表明该模型的表现优于 CN、AA 和 RA 算法[13]。N. Sett 等利用加性加权模型、最小流量模型和乘法模型在 10 个不同特征数据集上研究了连接权重对基于节点相似性的链接预测方法的影响,结果发现不同加权模型的表现会有所不同,该差异与所采用的预测方法和数据集有关[14]。K. K. Shang 等提出了直接连边预测算法,提高了演化网络上链路预测算法的精度,分析发现共同邻居数目对弱连接现象的形成起着重要作用[15]。B. Liu 等提出了一个结合零模型的通用框架,以量化拓扑结构、权重分布对加权网络中链路预测的影响[16]。K. J. Chen 等设计了一种名为 iBridge 的新链路预测方法,可以有效实现桥梁连接的识别,弥补了传统链路预测方法在预测弱连接方面的准确性不如强连接高的不足[17]。

综上已有研究,在取得成绩的同时,也存在一些不足之处:①从研究对象上看,对链路预测中的强弱连接效应问题的讨论大多集中在航空、合作、社交、神经等静态网络中,而鲜有针对学科引证知识扩散动态时变演化网络的研究。②链路预测指标的考察不够全面,引证关联强度对动态链路预测算法鲁棒性的影响有待明确。③对链路预测强弱连接现象与网络(尤其是有向网络)微观连边演化特性之间关系的把握不够清晰。因此,强弱连接对于学科引证知识扩散网络中动态链路预测的影响还有待进一步的探讨与思考。

采用不同算法指标进行知识扩散动态链路预测时,对于学科引证关联权重的强化或抑制起到了怎样的作用?不同权重的学科引证连边的缺失对预测算法精度有何影响?与邻居学科具有不同强弱引证关联模式的两学科之间究竟表现出怎样的连边倾向性?学科引证知识扩散行为演化规律对于动态链路预测算法指标的评价、设计与优化有何启示?

基于以上研究问题,本文依托内外协同(内因外象双重驱动)的思路理念,构建了一种外部网络结构调控与内部微观演化机理剖析相结合的动态链路预测强弱连接效应探测方法,从学科引证知识关联权重调节、连边失效触发以及强弱连接模体分析三个维度,对基于共同邻居相似性的学科引证知识扩散动态链路预测中的强弱连接效应问题进行了系统讨论。

2 研究设计

学科引证知识扩散动态链路预测强弱连接效应研究路线见图 1。该过程划分为两大模块三个阶段,其中两大模块是指外部网络结构调控和内部微观演化机理剖析模块,外部调控模块包括学科引证知识关联权重调节以及连边失效触发分析两个阶段,内部解析模块对应于强弱连接模体分析阶段,具体如下:

(1)基于权重调节的学科引证知识扩散动态链路预测强弱连接效应分析:在有向加权网络中的共同邻居相似性链路预测指标的基础上,引入参数 α 调节学科引证知识关联权重,测度参数调节下的强弱连接对预测学科引证知识扩散链路效果的影响。

(2)基于连边失效的学科引证知识扩散动态链路预测强弱连接效应分析:采用权重差异化连边(链路)失效策略,即按照连边权重大小,有选择地移除学科引证知识扩散网络中的连边,从缺失性视角揭示强弱连接对学科引证知识扩散动态链路预测算法维持其预测

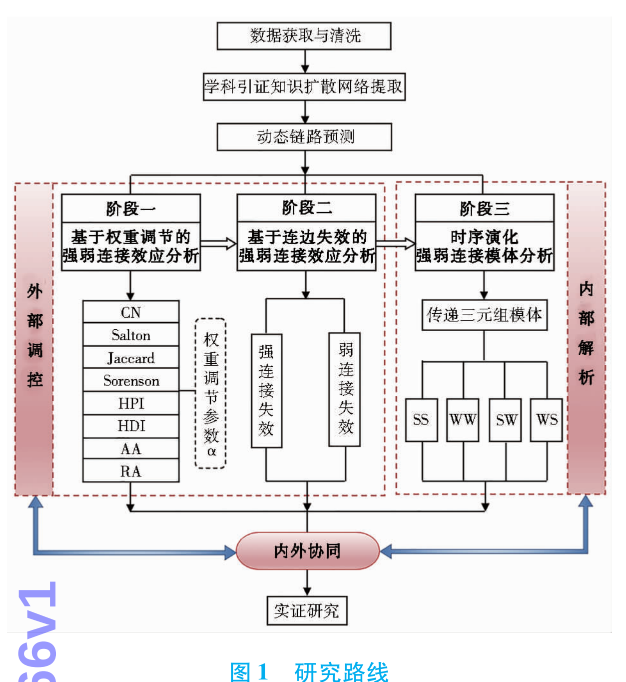


图 1 研究路线

效能的能力。

(3)学科引证知识扩散时序演化网络强弱连接模体分析:从学科引证知识扩散时序网络微观演化入手,进行三元组模体分析,计算强弱连接传递三元组模体连边比率,揭示不同权重影响下学科连边本质规律机理,为学科引证知识扩散动态链路预测算法指标的评价、设计与优化提供支撑。

2.1 学科引证知识扩散动态链路预测过程

在学科引证知识扩散网络 G 中,节点代表学科,连边代表学科间的引证知识连接。随着时间的推移,网络中的学科及其知识关联会发生动态变迁。将 t 时刻的网络 G_t 作为训练网络,给定一种链路预测算法,对 G_t 中每组学科对(x, y)赋予一个相似度值 S_{xy},该值代表了两学科间将来产生引证知识链路的潜力大小,从而形成预测网络。S_{xy} 越大,表示学科对之间借助文献引证发生知识扩散行为的可能性越高[18]。以 t+1 时刻的网络 G_{t+1} 作为测试网络,利用测试网络与预测网络的信息,结合学科引证知识扩散动态演化网络中的 AUC 评价指标[12,19],即测试网络 G_{t+1} 中新增加边(即在 G_{t+1} 中新衍生出的、原本在 G_t 中不存在的学科引证知识扩散连边)的相似度值比 G_{t+1} 中任意一条不存在边的相似度值高的概率,从整体上衡量链路预测算法的效能。为了确保评价结果的稳定性,以 1 年为间隔进行 q 次迭代预测(设初始时刻为 t_0,终止时刻为 t_q),分别计算历次预测的精度值 AUC_1、AUC_2、…、AUC_q,根据其数据分布特征选择合适的统计参量对指标效能进行综合考量[12]。

2.2 基于权重调节的学科引证知识扩散动态链路预测强弱连接效应分析

前期对不同链路预测指标在学科引证知识扩散演化网络中适用性的研究显示:LHN-Ⅰ指标的预测效果最差,稳定性也最低,不适于对学科引证知识扩散过程进行动态链路预测[12],因此,本文将其进行了排除。学科引证知识扩散网络中的含参数含权共同邻居相似性链路预测指标[2]如表 1 所示:

表 1 学科引证知识扩散网络中的含参数含权共同邻居相似性链路预测指标

相似性指标	公式
CN	$S_{xy}=\sum_{z\in\Gamma_{out}(x)\cap\Gamma_{in}(y)} w_{out}(x,z)^{\alpha}+w_{in}(z,y)^{\alpha}$
Salton	$S_{xy}=\sum_{z\in\Gamma_{out}(x)\cap\Gamma_{in}(y)} \frac{w_{out}(x,z)^{\alpha}+w_{in}(z,y)^{\alpha}}{\sqrt{S_{out}(x)\times S_{in}(y)}}$
Jaccard	$S_{xy}=\sum_{z\in\Gamma_{out}(x)\cap\Gamma_{in}(y)} \frac{w_{out}(x,z)^{\alpha}+w_{in}(z,y)^{\alpha}}{S_{out}(x)+S_{in}(y)-w(x,y)^{\alpha}}$
Sorenson	$S_{xy}=2\times\sum_{z\in\Gamma_{out}(x)\cap\Gamma_{in}(y)} \frac{w_{out}(x,z)^{\alpha}+w_{in}(z,y)^{\alpha}}{S_{out}(x)+S_{in}(y)}$
HPI	$S_{xy}=\sum_{z\in\Gamma_{out}(x)\cap\Gamma_{in}(y)} \frac{w_{out}(x,z)^{\alpha}+w_{in}(z,y)^{\alpha}}{\min\{S_{out}(x),S_{in}(y)\}}$
HDI	$S_{xy}=\sum_{z\in\Gamma_{out}(x)\cap\Gamma_{in}(y)} \frac{w_{out}(x,z)^{\alpha}+w_{in}(z,y)^{\alpha}}{\max\{S_{out}(x),S_{in}(y)\}}$
AA	$S_{xy}=\sum_{z\in\Gamma_{out}(x)\cap\Gamma_{in}(y)} \frac{w_{out}(x,z)^{\alpha}+w_{in}(z,y)^{\alpha}}{\lg(1+S_{out}(z))}$
RA	$S_{xy}=\sum_{z\in\Gamma_{out}(x)\cap\Gamma_{in}(y)} \frac{w_{out}(x,z)^{\alpha}+w_{in}(z,y)^{\alpha}}{S_{out}(z)}$

式中,对于学科 x、y 以及二者的共同邻居 z,定义 $\Gamma_{out}(x)$ 为学科 x 的出度邻居集,$\Gamma_{in}(y)$ 为学科 y 的入度邻居集;$w_{out}(x, z)$ 为学科 x 指向 z 的知识扩散连接的权重(即学科 z 引证 x 的频次),$w_{in}(z, y)$ 为学科 z 指向 y 的知识扩散连接的权重,w(x, y)为学科 x 指向 y 的知识扩散连接的权重;$S_{out}(x)=\sum_{z\in\Gamma_{out}(x)} w_{out}(x,z)^{\alpha}$ 表示学科 x 指向其邻居学科的知识扩散连接的调节权重之和,$S_{in}(y)=\sum_{z\in\Gamma_{in}(y)} w_{in}(z,y)^{\alpha}$ 表示学科 y 的邻居学科指向 y 的知识扩散连接的调节权重之和,$S_{out}(z)=\sum_{z\in\Gamma_{out}(x)\cap\Gamma_{in}(y)} w_{out}(z)^{\alpha}$ 表示学科 x、y 的共同邻居 z 指向 z 的邻居学科的知识扩散连接的调节权重之和。

当 α=0 时,即为无权学科引证知识扩散网络中的共同邻居相似性链路预测指标形式;当 α=1 时,即为加权网络中的链路预测指标形式。当 α<0 时,表示弱连接在链路预测中起到更加重要的作用;当 α>0 时,表示强连接在链路预测中起到更加重要的作用。α 的

绝对值越大，强弱连接的影响程度越显著。

2.3 基于连边失效的学科引证知识扩散动态链路预测强弱连接效应分析

学科引证知识扩散网络属于复杂网络的范畴，其异质性拓扑结构决定了网络中每条知识扩散链路的重要程度存在一定差异，一旦具有高价值（高知识负载、高枢纽性等）的链路发生断裂，大量与之相连的学科会失去连接方向，网络的结构和知识传输性能会发生改变[20]，从而导致依托于网络拓扑结构信息的学科引证知识扩散动态链路预测算法指标的性能受到影响，功能无法有效地发挥。本文采用权重差异化连边（链路）失效策略，从缺失性视角揭示强弱连接对学科引证知识扩散动态链路预测算法精度的影响，考察维持其预测效能的能力。

所谓连边失效是指按照一定的规则，有选择地移除学科引证知识扩散网络中的连边。权重差异化连边失效则特指按照连边权重大小，有选择地移除学科引证知识扩散网络中的连边。按照具有不同权重的连边失效顺序的不同，可划分为强连接失效和弱连接失效两类。

（1）强连接失效：按照学科间引证频次的高低（即连边权重）由大到小进行排序，由强到弱顺次移除连边。

（2）弱连接失效：按照学科间引证频次的高低（即连边权重）由小到大进行排序，由弱到强顺次移除连边。

选定一种链路预测算法，分别按照两种策略，移除网络中相同比例（10%、20%、30%、40%、50%、60%、70%、80%、90%、100%）的学科引证连边，形成学科知识扩散阻断网络 G_{ts} 和 G_{tw}，将其作为训练网络，以学科引证知识扩散完整网络 G_{t+1} 作为测试网络，计算 AUC 值，对比强弱连接缺失所造成的知识扩散动态链路预测性能差异。

2.4 学科引证知识扩散时序演化网络强弱连接模体分析

基于共同邻居相似性的动态链路预测算法主要依赖于网络中的拓扑结构信息，其预测效果的优劣在很大程度上取决于算法指标的设计能否抓住目标网络的结构特征及演化规律[21]。学科引证知识扩散网络随着时间的推移不断发生变化，这些变化主要体现在学科之间的引证情况以及学科知识扩散网络结构特征的改变。学科知识扩散网络结构特征主要受学科之间引证行为的影响，同时学科之间的引证连接会形成一些微观结构，如三个学科之间形成的团体结构，即三元组模体[8]。模体的不断变化、相互作用促使学科引证知识扩散网络中新连边的衍生。在学科引证知识扩散网络中，可能存在的三元组模体类型共 16 种[22]，如图 2 所示：

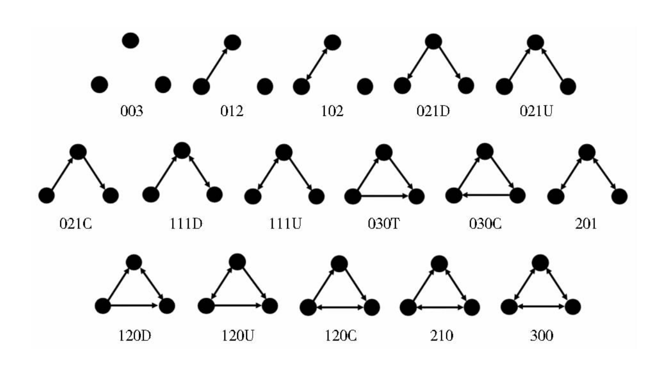


图 2　学科引证知识扩散网络中的 16 种三元组模体

每个类别由 3－4 个数字、字母表征，其中，第一个数字是指三元组模体中互惠对的个数；第二个数字是指不对称对的个数；第三个数字是指虚无对的个数；最后一个字母（如果需要的话）用于对相似的两个三元组模体进行区分："T"代表"传递"关系，"C"代表"循环"关系，"D"代表"向下"关系，"U"代表"向上"关系[23]。

对于由 x、y、z 三个学科构成的三元组模体，如果学科 z 引证了学科 x（即学科 x 的知识借助文献引证流向了学科 z），同时学科 y 引证了学科 z（即学科 z 的知识经由文献引证媒介扩散至学科 y），不管学科 x 和学科 y 之间是否存在引证知识关联，本文都将其统称为传递三元组模体。传递三元组模体是利用共同邻居相似性指标进行学科引证知识扩散动态链路预测的前提。根据学科间引证频次的大小，传递三元组模体中的连边可划分为 2 类：强连接和弱连接。根据学科引证知识扩散网络的连边权重分布情况，设定某一阈值 m，两学科间单向引证次数大于等于 m 的连边称为强连接，小于 m 的连边称为弱连接。学科引证知识扩散网络中，强弱连接传递三元组模体包括以下 4 种类型（ss、ww、sw、ws），见图 3。其中，学科 x 指向学科 z 以及学科 z 指向学科 y 的粗线表示强连接，细线表示弱连接；虚线表示可能存在连接，也可能不存在连接；学科 x 指向学科 y 的实线表示学科 x 的知识以文献引证为载体传播至学科 y，否则表示不存在连接。

以上 4 类传递三元组模体从强弱连接视角反映了邻居学科 z 与学科 x、学科 y 之间的引证知识关联模式：

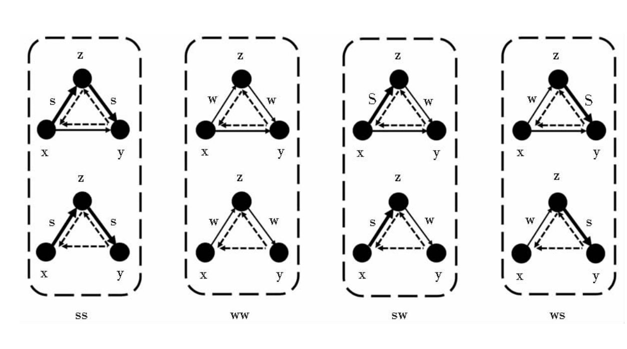


图 3 学科引证知识扩散网络中的强弱连接传递三元组模体

(1)ss:由强到强型,即学科 z 作为中介学科与知扩散源学科 x、知识扩散宿学科 y 都存在强引证关联;

(2)ww:由弱到弱型,即学科 z 作为中介学科与知扩散源学科 x、知识扩散宿学科 y 都存在弱引证关联;

(3)sw:由强到弱型,即知扩散源学科 x 与中介学科 z 存在强引证关联, 而中介学科 z 与知识扩散宿学科 y 存在弱引证关联;

(4)ws:由弱到强型,即知扩散源学科 x 与中介学科 z 存在弱引证关联,而中介学科 z 与知识扩散宿学科 y 存在强引证关联。

在学科引证知识扩散动态链路预测过程中,如果强连接起到更加重要的作用,那么对于由学科 x、学科 z 以及学科 z、学科 y 构成的两条强连接,即 ss 型传递三元组模体来说,学科 x 和学科 y 相连接的概率就较大。同理,如果弱连接起到更加重要的作用,那么对于由学科 x、学科 z 以及学科 z、学科 y 构成的两条弱连接,即 ww 型传递三元组模体来说,学科 x 和学科 y 相连接的概率就较大。基于此,以 Pss、Pww、Psw、Pws 分别表示不同强弱连接传递三元组模体类型中具有共同邻居的两学科存在引证知识扩散关联的比率,进而衡量不同权重影响下学科连边的倾向性,其含义及意义如表 2 所示:

表 2 学科引证知识扩散网络强弱连接传递三元组模体连边比率

指标	含义	意义
Pss	学科 z 与知扩散源学科 x、知识扩散宿学科 y 都存在强引证关联时,学科 x 直接向学科 y 以文献引证的形式溢出知识的概率	反映了两学科通过包含 2 条强连接的路径相连通时,其直接存在引证知识扩散关系的可能性大小。该值越大,说明强连接在学科引证知识扩散网络演化及动态链路预测过程中起到更加重要的作用
Pww	学科 z 与知扩散源学科 x、知识扩散宿学科 y 都存在弱引证关联时,学科 x 直接向学科 y 以文献引证的形式溢出知识的概率	反映了两学科通过包含 2 条弱连接的路径相连通时,其直接存在引证知识扩散关系的可能性大小。该值越大,说明弱连接在学科引证知识扩散网络演化及动态链路预测过程中起到更加重要的作用
Psw	学科 z 与知扩散源学科 x 存在强引证关联,而与知识扩散宿学科 y 存在弱引证关联时,学科 x 直接向学科 y 以文献引证的形式溢出知识的概率	反映了两学科通过包含 1 条强连接和 1 条弱连接的路径相连通时,其直接存在引证知识扩散关系的可能性大小。该值越大,说明知扩散源学科的溢出辐射能力高于知识宿学科的吸纳融合能力时,二者之间产生直接知识转移行为的倾向性越明显
Pws	学科 z 与知扩散源学科 x 存在弱引证关联,而与知识扩散宿学科 y 存在强引证关联时,学科 x 直接向学科 y 以文献引证的形式溢出知识的概率	反映了两学科通过包含 1 条弱连接和 1 条强连接的路径相连通时,其直接存在引证知识扩散关系的可能性大小。该值越大,说明知识宿学科的吸纳融合能力高于知扩散源学科的溢出辐射能力时,二者之间产生直接知识转移行为的倾向性越明显

3 研究对象的选择与数据获取

根据前期对于学科引证知识扩散动态链路预测的研究[12],本文选取预测效果稳定性较高的 2006–2016 年社会网络领域学科引证知识扩散网络为研究对象,探索学科引证知识强弱连接对于动态链路预测过程的影响。

在 Web of Science 中,以“TS =“social network *””为检索策略,收集 2006–2016 年 SCI-EXPANDED、SSCI 数据库中,类型为“articles”的文献,去重、清洗后得到文献 25 539 篇。根据期刊引证报告(JCR)的期刊所属学科类别对照表,在期刊引证数据的基础上,抽取历年学科引证知识扩散时序演化网络。

4 数据处理与结果分析

4.1 基于权重调节的学科引证知识扩散动态链路预测强弱连接效应分析

含权学科引证知识扩散动态链路预测共同邻居相似性指标的 AUC 值随参数 α 的变化情况如图 4 所示。文中所涉及的 AUC 值均为利用 2006–2016 年的时序演化网络,以 1 年为间隔进行 10 次迭代动态链路预测所得预测精度的平均值[12]。

由图 4 可知,除 HDI 指标外,随着参数 α 逐渐增大,其他 7 项指标的预测效能都呈现先上升后下降的趋势,但变动幅度不尽相同。其中,CN、AA、RA 指标精度先是迅速增加,当达到峰值后,缓慢下降;Salton、HPI 指标精度在攀升至顶峰后,迅速下滑;Jaccard、Sorenson 指标精度的演变形态近于一致,都表现为先微弱上升,

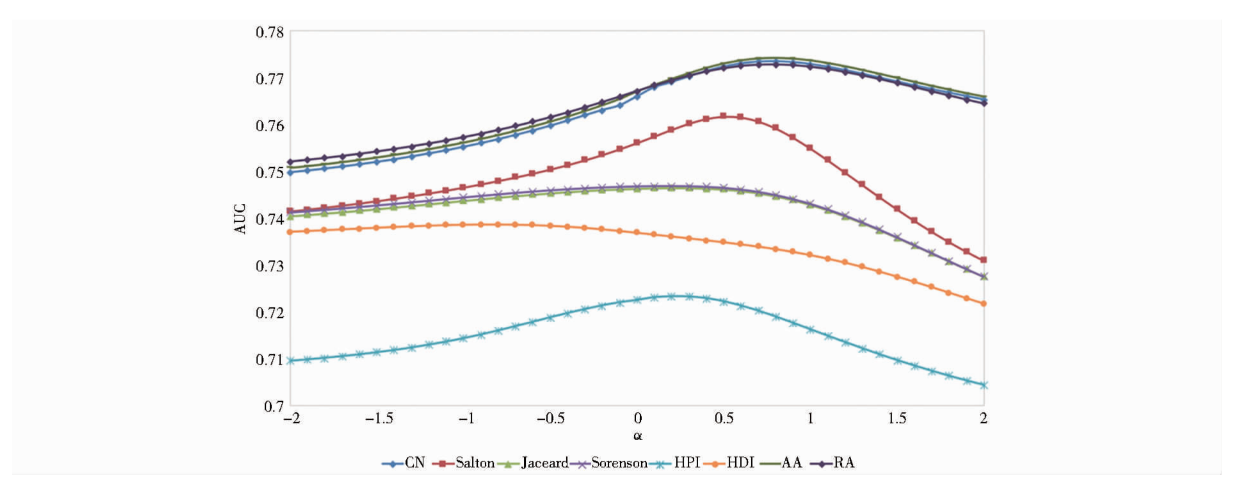


图 4 含权学科引证知识扩散动态链路预测共同邻居相似性指标的 AUC 值随参数 α 的变化

后逐渐下降的态势。当 α 小于 0 时，HDI 指标的预测表现较为平稳；当 α 大于 0 后，则呈现波动下滑态势。

表 3 为无权、加权以及含权最优参数(即使得预测准确度达到最高值的参数 α)下学科引证知识扩散动态链路预测共同邻居相似性指标的 AUC 值。

表 3 无权、加权以及含权最优参数下学科引证知识扩散动态链路预测共同邻居相似性指标的 AUC 值

共同邻居相似性指标	CN	Salton	Jaccard	Sorenson	HPI	HDI	AA	RA
无权	0.765 82	0.755 94	0.746 25	0.746 66	0.722 37	0.736 73	0.766 88	0.767 01
加权	0.772 71	0.754 82	0.742 81	0.742 95	0.716 13	0.731 95	0.773 51	0.772 22
最优参数 α	0.773 30	0.761 59	0.746 39	0.746 74	0.723 15	0.738 49	0.774 03	0.772 70

由表 3 可知，对于学科引证知识扩散时序演化网络来说，8 项指标在最优参数调节下的预测表现较单纯无权和加权指标均有不同程度的提升。其中，AA 指标在最优参数 α 下的预测效果最好(其预测精度均值达 0.77403)，CN、RA 指标的表现次之，HPI 指标的预测精度最优均值仅为 0.72315，居于末位。

表 4 为不同共同邻居相似性指标下的链路预测最优参数 α。

表 4 不同共同邻居相似性指标下学科引证知识扩散动态链路预测最优参数 α

共同邻居相似性指标	CN	Salton	Jaccard	Sorenson	HPI	HDI	AA	RA
最优参数 α	0.8	0.5	0.2	0.2	0.2	-0.8	0.8	0.8

表 4 中，HDI 指标的最优参数为负值，说明利用其进行动态链路预测时，如果提升弱知识引证连边的权重，降低强连边的权重，会得到更佳预测表现，即弱连接扮演着更加重要的角色。

其余 7 项指标的最优参数都为正值，说明利用这些指标进行动态链路预测时，强连接起着更为重要的作用。然而，其最优参数都小于 1，说明当学科间的引证关联强度发生适度衰减时($0 < \alpha < 1$)，往往能取得最佳动态链路预测效果，也就是说，虽然强连接更重要但是并未达到其权重显示的重要程度。CN、AA、RA 指标的最优参数值相对于其他 4 项指标较大，说明三者的强连接效应影响更加显著。

对于不同的共同邻居相似性指标来说，强弱连接关系对学科引证知识扩散动态链路预测过程的影响程度存在一定差异。但是，综合来看，学科引证知识扩散时序演化网络中的强连接效应更为明显，即权重较大的强连接对于预测准确度的提升贡献更大。

4.2 基于连边失效的学科引证知识扩散动态链路预测强弱连接效应分析

利用 Matlab 模拟不同连边失效策略下，不同相似性算法指标在学科引证知识扩散时序演化网络中的动态链路预测性能变化，结果如图 5 所示(横坐标为按照学科引证连边权重大小，由强到弱/由弱到强顺次移除连边的比例；纵坐标为利用移除连边后的网络进行迭代动态链路预测时的 AUC 值)。

由图 5 可看出，随着学科引证连边的减少，各项指标的预测精度出现不同程度的损失，当学科间的引证关联全部消失(即移除边数比例达到 100%)时，动态链路预测 AUC 值下降至 0.5，与所有预测连边的相似度值都是随机产生的情形相当。对于 8 项指标来说，

图书情报工作
第 65 卷 第 13 期 2021 年 7 月

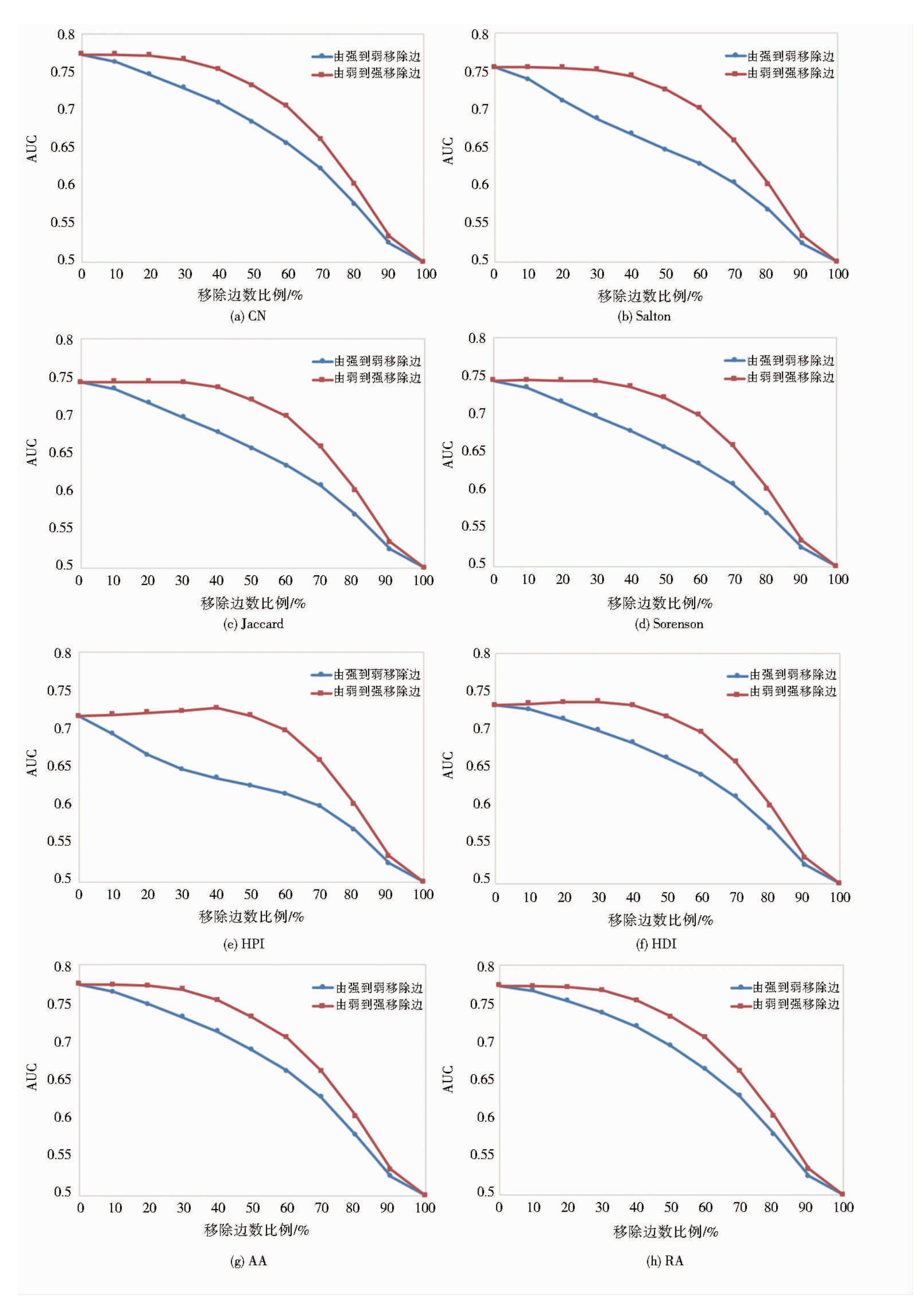


图 5　不同连边失效策略下学科引证知识扩散动态链路预测性能变化

移除弱连接时的预测精度下降速度都较为平缓，弱连接移除量达到40%左右时，指标预测 AUC 值才开始出现较为明显的下滑，说明弱连接失效对于学科引证知识扩散动态链路预测效能的影响较小，即学科引证知识扩散时序演化网络对于弱连接失效具有较强的鲁棒性，其弱连接容错能力较大。而强连接失效策略下，预测精度的变化较为急促，说明强连接对于学科引证知识扩散动态链路预测性能维持度的作用更加明显，即网络的强连接抗攻击能力较小。两种连边失效策略下，Salton、HPI 指标的动态链路预测性能变化曲线差异最大，Jaccard、Sorenson、HDI 指标次之，CN、AA、RA 指标的差异最小，说明不同指标对于强弱连接（尤其是强连接）缺失的敏感性存在一定差异，预测精度衰减速度越快，其敏感性越高。

值得注意的是，前述研究发现 HDI 指标的最优权重调节参数为负值，即利用其进行学科引证知识扩散

动态链路预测时,弱连接起到更加重要的作用。而权重差异化连边失效策略下,对于 HDI 指标来说,强连接对于其预测效能维持能力的影响更加明显。二者似乎存在矛盾之处,究其原因在于,移除连边策略在改变强弱连接权重的同时,学科间共同邻居的数目也会发生变化,模体结构分布亦随之改变,因此同一指标的预测精度变化趋向也可能会产生一定的差异。

4.3 学科引证知识扩散时序演化网络强弱连接模体分析

学科引证知识扩散时序演化网络中,16 种三元组模体类型的相对比率如图 6 所示:

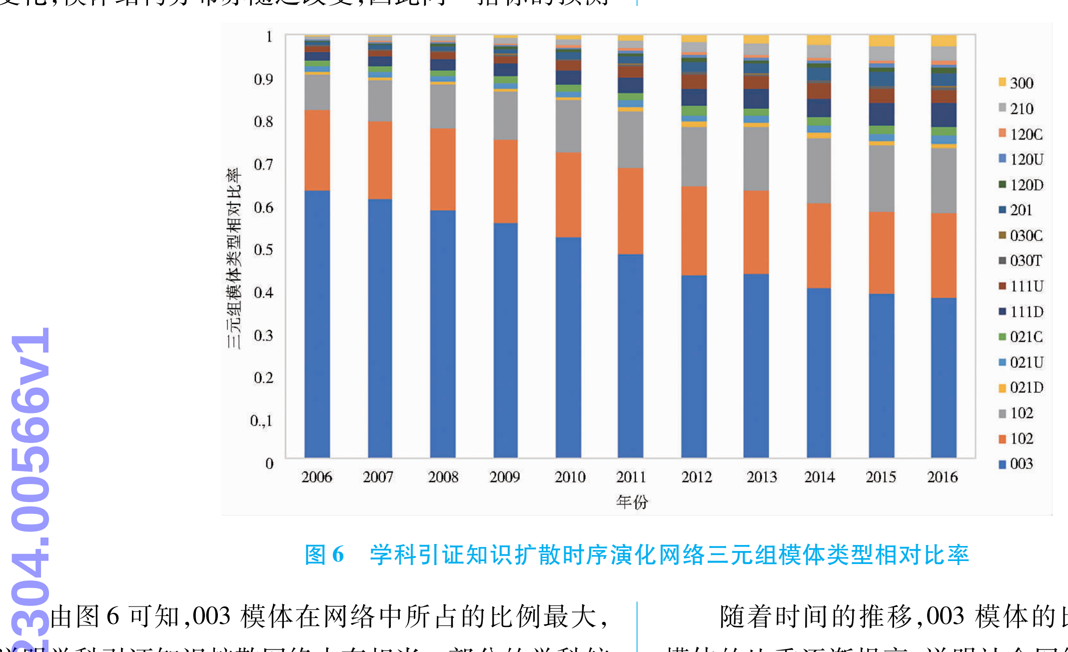


图 6 学科引证知识扩散时序演化网络三元组模体类型相对比率

由图 6 可知,003 模体在网络中所占的比例最大,说明学科引证知识扩散网络中有相当一部分的学科较为孤立,与其他学科之间的引证知识关联比较稀疏。012 模体和 102 模体的比重次之,说明学科之间的二元知识扩散直接连接较三方联系更为普遍,且两个学科之间的单向知识流动比双向互惠融合特性更加显著。另外,网络中 021U 模体数量多于 021D 模体,说明对于社会网络领域学科来说,当其邻居学科之间不存在直接引证关系时,其共引知识汇聚的能力要大于耦合知识渗透能力。

随着时间的推移,003 模体的比重越来越小,102 模体的比重逐渐提高,说明社会网络领域学科之间的引证知识交流(尤其是双向知识传播)越来越密切,网络连通性稳步提升,愈加利于知识在学科中的横向拓展及纵向渗透。

然而,以上类型的模体都不具备传递性,即不属于传递三元组模体。传递三元组模体是利用共同邻居相似性指标进行学科引证知识扩散动态链路预测的前提。学科引证知识扩散时序演化网络中,11 种传递三元组模体类型的相对比率如图 7 所示:

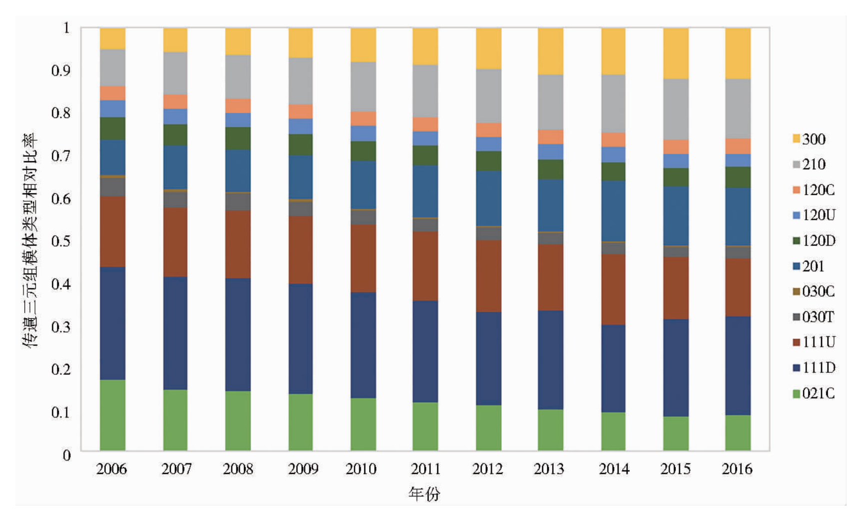


图 7 学科引证知识扩散时序演化网络传递三元组模体类型相对比率

由图 7 可知，在具有传递性的三元组模体中，021C、111D、111U 和 201 模体占有较大的比重，说明相当一部分与中介学科都存在知识引证关系的其他两学科之间并不直接连通。随着时间的推移，210、300 模体的数量不断增加，学科间的双向知识交流关联愈加富集，互惠连通率逐渐提升。

为进一步从引证知识关联强度的视角，明晰学科引证知识扩散网络的时序演化本质规律，以 Pss、Pww、Psw、Pws 分别表示不同强弱连接传递三元组模体类型中具有共同邻居的两学科存在引证知识扩散关联的比率，进而衡量不同权重影响下学科连边的倾向性，为学科引证知识扩散动态链路预测算法指标的设计优化提供支撑。2006 – 2016 年间共涉及 11 个学科引证知识扩散时间切片网络，笔者采用间隔抽样的方式，抽取了 6 个子网络进行分析，强弱连接传递三元组模体连边比率如图 8 所示。其中，横坐标表示该年学科引证知识扩散网络中连边权重由小到大排列的序号，具体来说，先将学科之间互不重复的引证连边权重值由小到大排列（最小的边权重为 0，即不存在学科引证关联），然后用自然数（0、1、2、……）依次给这些连边权重赋予等级序号，该序号代表对应连边权重大小在边权集中所处的相对位置；纵坐标表示以不同连边权重作为强弱连接划分阈值时，学科引证知识扩散网络中由强到强（ss）、由弱到弱（ww）、由强到弱（sw）以及由弱到强（ws）四种强弱连接传递三元组模体的连边比率。

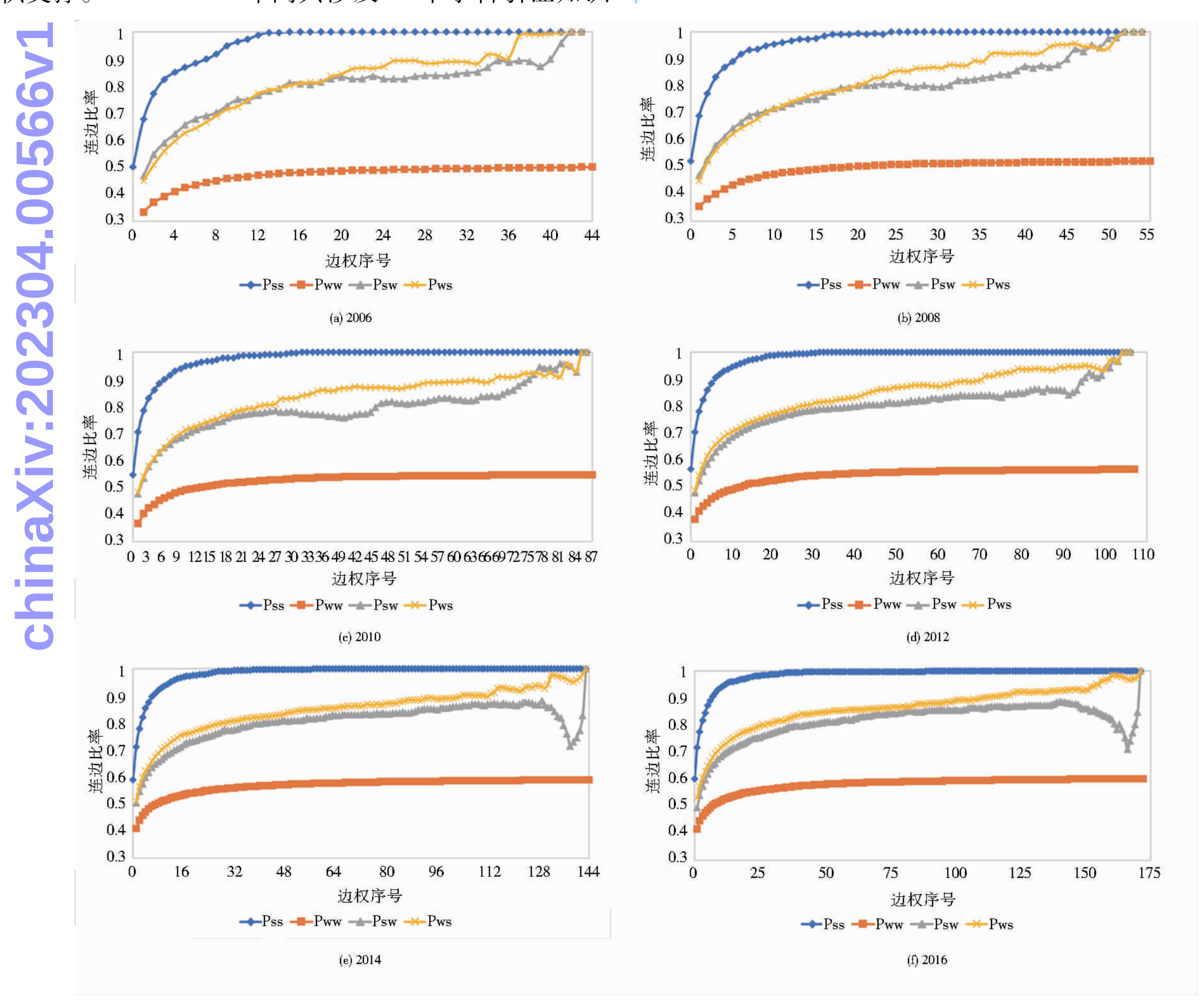


图 8　学科引证知识扩散时序演化网络强弱连接传递三元组模体连边比率

由图 8 可知，在社会网络领域学科引证知识扩散网络中，两学科通过包含 2 条强连接的路径相连通时，其直接存在引证知识扩散关系的可能性（Pss）要远大于通过包含 2 条弱连接的路径相连通时，其直接存在引证知识扩散关系的可能性（Pww）。尤其是当邻居学科 z 与知扩散源学科 x、知识扩散宿学科 y 都存在非常强的引证关联（即连边权重都很大）时，学科 x 直接向学科 y 以文献引证的形式溢出知识的概率可达到 100%。因此，从整体上看，强连接比弱连接更能够促进学科间引证知识关联链路的产生，在学科引证知识

岳增慧，许海云，赵敏. 强弱连接对学科引证知识扩散动态链路预测的影响研究[J]. 图书情报工作,2021,65(13):66 - 76.

扩散网络演化及动态链路预测过程中起到更加重要的作用。

另外，自 2012 年以来，当两学科通过一强一弱 2 条路径相连通时，若知识宿学科的吸纳融合能力高于知识扩散源学科的溢出辐射能力，则二者之间产生直接知识转移行为的概率更会高（Pws > Psw），且该演化特性与强弱连接切分阈值的大小无关。

5 结论

本文依托内外协同的思路理念，构建了一种外部网络结构调控与内部微观演化机理剖析相结合的动态链路预测强弱连接效应探测方法，从学科引证知识关联权重调节、连边失效触发以及强弱连接模体分析三个维度，对学科引证知识扩散动态链路预测中的强弱连接效应问题进行了深入探索与系统讨论，为强弱连接理论在有向加权动态网络链路预测中的应用拓展提供了支撑，丰富与完善了动态链路预测中强弱连接效应探测方法体系，为学科引证知识扩散行为微观演化规律的揭示以及动态链路预测算法指标的评价、设计与优化提供了可资借鉴的理论和实践参考。主要得到以下结论：

（1）对于不同的共同邻居相似性指标来说，强弱连接关系对学科引证知识扩散动态链路预测过程的影响程度存在一定差异。链路预测"强弱连接效应"是针对算法指标而言的，并不完全代表网络本身的连边演化特性。综合来看，权重较大的强连接对于预测准确度的提升贡献更大。

（2）从缺失性视角来看，强连接对于学科引证知识扩散动态链路预测效能维持能力的作用更加明显；学科引证知识扩散时序演化网络对于弱连接失效具有较强的鲁棒性，其弱连接容错能力较大。不同指标对于强弱连接（尤其是强连接）缺失的敏感性也不尽相同。另外，动态链路预测中的强弱连接现象不仅与学科引证关联权重有关，还会受到共同邻居数目以及网络微观模体结构的影响。

（3）随着时间的推移，社会网络领域学科之间的引证知识交流（尤其是双向知识传播）越来越密切，网络连通性稳步提升，为知识在学科中的横向拓展及纵向渗透提供了有利条件。两学科与邻居学科之间的引证关联强度越大，则这两个学科之间直接发生知识转移的概率越高。同时，知识宿学科的吸纳融合能力相对于知识源学科的溢出辐射能力来说，在新连边的衍生过程中的主导地位更加突出。

（4）大部分基于共同邻居相似性的链路预测指标的设计理念基本符合学科引证知识扩散网络的连边演化规律，但对于强弱连接权重因素的考量仍有待进一步完善。从整体上看，强连接在学科引证知识扩散网络演化及动态链路预测过程中扮演着更加重要的角色。因此，突出大权重强连接在预测指标设计中的作用，是提升学科引证知识扩散动态链路预测性能的有效途径。

此外，本文仅对社会网络领域范围内基于共同邻居相似性的学科引证知识扩散动态链路预测过程中的强弱连接效应问题进行了分析，结论的可推广性有待进一步验证。虽然研究发现强连接占主导地位，然而弱连接作为一种特有的学科交互模式在知识传播演化过程中也发挥了不可忽视的作用。具有强连接的学科间往往具有较大的知识同质性，学科吸收、融合和传播同质性知识，一定程度上有助于学科在现有基础上进行理论、技术、方法等方面的常规性发展与渐进性创新；而具有弱连接关系的学科间知识资源异质化特征明显，此类知识的交叉融合有利于颠覆性技术生长点的衍生，从而推动学科取得突破性创新，实现飞跃式发展。强连接和弱连接相辅相成，如何在学科知识扩散动态链路预测过程中实现强弱连接有机结合，是有待深入研究的重要问题之一。

下一步将在深入挖掘并准确把握学科引证知识扩散行为本质机理的基础上，从模体演化微观视角出发，合理利用强弱连接权重信息，同时从中观层面结合学科社区属性信息，优化学科引证知识扩散时序动态链路预测方法，提升预测效能。

参考文献：

[1] GRANOVETTER M S. The strength of weak ties[J]. American journal of sociology, 1973, 78(6): 1360 - 1380.
[2] 吕琳媛，周涛. 链路预测[M]. 北京：高等教育出版社，2013.
[3] 张淑谦，徐顺治，李登科. 社会网络理论及其研究述评[J]. 智库时代，2019(37):264,268.
[4] 滕岩松. 基于众包平台 LineMe 的社会网络演化与透明性研究[D]. 上海：上海交通大学，2018.
[5] HANSEN M T. The search-transfer problem: the role of weak ties in sharing knowledge across organization subunits[J]. Administrative science quarterly, 1999, 44(1): 82 - 111.
[6] MARTIN R. Strong ties, weak ties and islands: structural and cultural predictors of organizational innovation[J]. Industrial & corporate change, 2002, 11(3): 427 - 449.
[7] FRIEDKIN N E. Information flow through strong and weak ties in intraorganizational social networks[J]. Social networks, 1982, 3(4):273 - 285.
[8] MILO R, SHEN-ORR S, ITZKOVITZ S, et al. Network motifs: simple building blocks of complex networks[J]. Science, 2002, 298(5594): 824 - 827.
[9] 刘书新. 基于非对称神经网络的跨网络社交关系预测研究[D]. 重庆：重庆邮电大学，2019.
[10] MURATA T, MORIYASU S. Link prediction of social networks

based on weighted proximity measure [C]. Proceedings of IEEE/WC/ACM international conference on web intelligence. New York: ACM Press, 2007.
[11] LÜ L Y, ZHOU T. Link prediction in weighted networks: the role of weak ties [J]. EPL (Europhysics letters), 2010, 89(1): 18001.
[12] 岳增慧,许海云,王倩飞. 基于局部信息相似性的学科引证知识扩散动态链路预测研究[J]. 情报理论与实践,2020,43(2):84-91,99.
[13] LIU H, HU Z, HADDADI H, et al. Hidden link prediction based on node centrality and weak ties [J]. EPL (Europhysics letters), 2013, 101(1):18004.
[14] SETT N, RANBIR SINGH S, NANDI S. Influence of edge weight on node proximity based link prediction methods: an empirical analysis [J]. Neurocomputing, 2016, 172:71-83.
[15] SHANG K K, SMALL M, XU X K, et al. The role of direct links for link prediction in evolving networks [J]. EPL (Europhysics letters), 2017, 117(2):28002.
[16] LIU B, XU S, LI T, et al. Quantifying the effects of topology and weight for link prediction in weighted complex networks [J]. Entropy, 2018, 20(5):363.
[17] CHEN K J, ZHANG P, YANG Z N, et al. iBridge: inferring bridge links that diffuse information across communities [J]. Knowledge-based systems, 2020, 192:105249.
[18] 吕琳媛. 复杂网络链路预测[J]. 电子科技大学学报,2010,39(5):651-661.
[19] HANLEY J A, MCNEIL B J. The meaning and use of the area under a receiver operating characteristic (ROC) curve [J]. Radiology, 1982, 143(1): 29-36.
[20] 岳增慧,方曙. 科研合作网络弹性研究与实证[J]. 图书情报工作,2013,57(11):86-89,95.
[21] 张斌,马费成. 科学知识网络中的链路预测研究述评[J]. 中国图书馆学报,2015,41(217):99-113.
[22] HOLLAND P W, LEINHARDT S. A method for detecting structure in sociometric data [J]. American journal of sociology, 1970, 76(3): 492-513.
[23] 刘军. 整体网分析:UCINET 软件实用指南(第三版)[M]. 上海:格致出版社,2019.

作者贡献说明:
岳增慧:研究设计、数据分析和论文撰写;
许海云:数据处理与分析;
赵敏:论文修订。

Influence of Strong and Weak Ties on Dynamic Link Prediction of Knowledge Diffusion in Disciplinary Citation Networks

Yue Zenghui[1] Xu Haiyun[2] Zhao Min[3]
[1] School of Medical Information Engineering, Jining Medical University, Rizhao 276826
[2] Business School, Shandong University of Technology, Zibo 255000
[3] Jining Medical University, Rizhao 276826

Abstract: [**Purpose/significance**] The strong and weak ties is one of the important factors that affect the dynamic link prediction of knowledge diffusion in disciplinary citation networks. The strong and weak citation ties in diffusion of disciplinary knowledge jointly promote knowledge exchange, integration and innovation among disciplines with coordination and mutual effect. The exploration of strong and weak ties in the dynamic link prediction of knowledge diffusion in disciplinary citation networks can provide theoretical and practical references for expanding the application of the strong and weak ties theory, revealing the micro-evolution law of knowledge diffusion behavior of disciplinary citation, and evaluating, designing and optimizing dynamic link prediction algorithm indicators. [**Method/process**] In this paper, on the basis of synergistic idea, the method of detecting strong and weak ties in the dynamic link prediction was constructed by controlling external structure and analyzing internal evolution mechanism of the networks. To be specific, the influence of strong and weak ties on the dynamic link prediction of knowledge diffusion in disciplinary citation networks based on common neighbor similarity was discussed from three perspectives of adjustment of knowledge connection weight in disciplinary citation networks, link failure triggering and motif analysis. [**Result/conclusion**] The research has shown that, firstly, strong ties play a more important role in evolution of knowledge diffusion in disciplinary citation networks and dynamic link prediction; secondly, strong and weak ties effect in link prediction is not only related to connection weight of disciplinary citation networks, but affected by the number of common neighbor and micro motif structure; thirdly, compared with the spillover ability of knowledge from source discipline, the absorbing ability of knowledge in destination discipline has a more prominent impact on the process of developing new knowledge links.

Keywords: knowledge diffusion in disciplinary citation networks　strong ties and weak ties　weight adjustment　link failure　triad motif　dynamic link prediction